Motivation and Objectives

- Trade-off such as “how much is enough” is challenging, esp. in exploratory testing
  - Insufficient testing can lead to unsatisfying product quality, while
  - Excessive testing can result in potential schedule delays and low cost-effectiveness.

- Objective
  - To enable actionable, value-driven decision making on resource allocation and utilization faced by testing managers.

- Plan for the close of tasks largely based on expert knowledge
- Employ a fixed period or a fixed number of participants
Project Overview

• MET project
  — Investigates, characterizes distributing testing processes
  — Develops a set of Machine Learning-based approaches to support efficient testing management across distributed testing teams

• MET consists of
  — A testing measurement model
    o For characterizing the representative contextual factors of a testing process
  — A in-process team formation model
    o Matching, learning, ranking, and dynamically tuning the configuration of distributed testing teams to maximize testing adequacy
    o Leveraging natural language processing (NLP) and learning-to-rank algorithms
  — A early completion detection model
    o Monitoring, aggregating testing reports, predicting total defects, and automate testing completion detection
Pilot Study – Analysis Results

- A pilot study: 636 real-world mobile application testing projects.

- Observation: Excessive engagement leading to waste due to duplicated effort.
  - Average wasteful spending over 636 projects: 32%

**Bug Arrival Patterns:**
1) Large variation in bug detection speed and cost;
2) Decreasing bug detection rates over time;
3) Plateau effect of bug arrival curves.
Evaluation: Baidu Crowd Test dataset;
1. Automated close management: 30% cost reduction.
2. Trade-off analysis support.

Testing Measurement Model (TMM)

• Two sub-models
  
  — **Process context sub-model**: in-process progress-oriented information
    
    o **Testing requirement matrix**: represent the task’s requirements in the vector space of descriptive terms from testing requirements.
    
    o **Test adequacy**: To measure the testing progress regarding to what degree each descriptive term has been tested.
  
  — **Resource context sub-model**: characteristics of distributed testing teams
    
    o **Activeness**: a set of metrics to characterize a team’s activeness over certain period
    
    o **Preference**: distribution of term intensity from a testing team’s historical reports
    
    o **Expertise**: distribution of term intensity from a testing team’s unique bug reports
    
    o **Devices**: environmental factors such as testing devices, equipment, etc.
**iRec: In-Process Team Recommendation**

**Evaluation:** Baidu Crowd Test dataset; Reduction of the NYWs by 50% - 58% ; Reduction of cost by about 10% on median

Work in progress: iSENSE 2.0

• Extending iSENSE with new components
  — Automated Duplicate Tagger
    o Analyzing the duplicate status of received crowd reports leveraging on semantic analysis
  — Coverage-based Sanity Checker
    o Reinforce the stability and performance of close prediction.

Note that, the components with Green background are newly proposed in iSENSE2.0.
Next Steps

• Elaborate the testing measurement model (TMM)
  — To address specific needs for characterizing DoD task/process/resource context
  — Refine core underlying metric: Testing Process Adequacy

• Integrate TMM with iRec and iSENSE 2.0

• Empirical validation of existing models on cross-platform datasets
  — In-process team recommendation
  — Early completion detection

• Seeking collaboration in further evaluation in DoD testing projects.
Thank you!

Contact:
ye.yang@stevens.edu